scientific reports

Check for updates

OPEN

Gastrointestinal tract disease classification from wireless capsule endoscopy images based on deep learning information fusion and Newton Raphson controlled marine predator algorithm

Saddaf Rubab[1]✉, Muhammad Jamshed[2], Muhammad Attique Khan[2], Nouf Abdullah Almujally[3], Robertas Damaševičius[4], Amir Hussain[5], Neunggyu Han[6] & Yunyoung Nam[6]✉

Worldwide, cancer is one of the leading causes of death in humans. Interobserver variability and specialized experience are key factors in diagnosing gastrointestinal tract (GIT) abnormalities using endoscopic procedures. Due to this diversity, small lesions may go unnoticed, leading to a delay in early diagnosis. Therefore, it is essential to design a computer-aided diagnosis (CAD) system for the detection and classification of GIT diseases at the early stages. This paper proposes a CAD system that combines the feature fusion of modified deep learning models with optimal feature selection. Three publicly available datasets, including Kvasir V1, Kvasir V2, and Hyperkvasir, are utilized in the experimental process. In the proposed method, a contrast enhancement step is performed using the fusion of the top-bottom filtering technique. In the next step, two deep learning models (ResNet18 and ResNet50) are modified with a new layer called entropic field propagation (EFP). The pooling layers are replaced with EFP layers in both models, which are then trained on the selected datasets. In the testing process, trained models are employed, and features are extracted from the deeper layers, which are further refined using the Newton-Raphson Marine Predator Optimization (NRMPO) algorithm. The selected features from both models are finally fused using a novel mean threshold-based fusion approach and passed to machine learning classifiers. The proposed CAD system achieved accuracies of 99.0, 89.6, and 82.7% for Kvasir V1, Kvasir V2, and HyperKvasir, respectively. A detailed ablation study is also conducted for the middle steps that validate these reported accuracies. C*onclusion*: A comparison is performed with state-of-the-art (SOTA) techniques, showing that the proposed method achieves improved accuracy and precision rates.

Keywords Stomach cancer, Wireless capsule endoscopy, Deep learning, Fusion, Optimization, Classification

Identification of gastrointestinal tract (GIT) infections is a complex medical research problem caused by the production of abnormal cells that form a mass in part of the stomach[1]. Environmental factors, including a person's eating habits and community behaviors, are known to cause gastric cancer[2]. The symptoms of gastrointestinal tract infections include ulcers, bleeding, and polyps; however, the number of stomach cancer cases is increasing, with well-defined global variations[3]. The anticipated instances of stomach cancer in the US

[1]Department of Computer Engineering, College of Computing and Informatics, University of Sharjah, Sharjah 27272, United Arab Emirates. [2]Center of AI, Prince Mohammad bin Fahd University, Al-Khobar, Saudi Arabia. [3]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O.Box 84428, 11671 Riyadh, Saudi Arabia. [4]Kaunas University of Technology, 44249 Kaunas, Lithuania. [5]School of Computing, Edinburgh Napier University, Edinburgh, UK. [6]Department of ICT Convergence, Soonchunhyang University, Asan 31538, Republic of Korea. ✉email: srubab@sharjah.ac.ae; ynam@sch.ac.kr

since 2020 were 27,600, of which 16,980 were male and 10,620 were female. The estimated death rate was 11,010, of which 6,650 were male and 4,360 were female[4]. New cases are added to the general number of bowel infections each year[5]. In 2023, there were 26,500 recorded cases of stomach cancer, while there was an 11,130 death rate. Due to the high death rate, early diagnosis is challenging[6]. One of the most commonly used diagnostic methods for gastrointestinal illnesses is conventional gastroscopy, which may accurately identify tumors[7]. Conventional gastroscopy, on the other hand, is complex and intrusive, and is inefficient at detecting abnormalities in the small intestine[8].

Recently, Wireless Capsule Endoscopy (WCE) technology has gained popularity, as it enables medical professionals to examine the small intestine, a difficult-to-access location. The patient consumes the capsule in this method, and the capsule endoscopy camera records hundreds of color images as it travels through the digestive system[9]. Because this diagnosis has the obvious drawbacks of being costly and time-consuming, it also incurs high diagnostic costs due to the involvement of a medical specialist, specifically a gastroenterologist. Researchers have proposed automated techniques, such as Computer Vision (CV)-based methods, which many studies have adopted to automatically identify different significant gastric tumors in WCE images. Many computerized methods for diagnosing and treating medical conditions have emerged in recent years[10].

Recently, various computerized methods for diagnosing and classifying diseases have been developed, including deep learning (DL), which has significantly enhanced medical image processing[11,12]. Preprocessing, which often precedes this procedure, is a crucial step in ensuring high process accuracy[13]. At this stage, images are enhanced before proceeding to augmentation, feature extraction, and classification[14]. The feature extraction strategy may be helpful when we have a vast data collection and need to minimize waste without losing any crucial or valuable information. With the use of feature extraction, the number of duplicated data points in the data collection is reduced[15]. To select the best features, feature selection is employed. Feature selection, as an essential phase in deep learning, is one of the most widely used and significant data preparation techniques that has been developed[16]. This procedure accelerates data mining algorithms, enhances prediction accuracy, and makes data more understandable. Approaches based on deep learning have successfully transitioned into the medical imaging industry and excelled in the CV field[17,18]. Through the combination of lower-level convolutional layers with higher-level features, deep learning enables the extraction of features in a hierarchical structure. Deep convolutional neural network (DCNN) models that have already been trained have been proposed by several researchers in the CV community, including the well-known AlexNet, ResNet-18, VGG-19, and GoogleNet[19]. Many researchers have employed these models for transfer learning in medical imaging, as they were developed using the ImageNet dataset. According to earlier research[20], it has been proven that enhancing the most favorable features leads to more correct classification. Feature optimization aims to reduce the learning time of new features and enhance the system's precision[21]. We employed Marine Predator algorithms to optimize the feature selection problem and fused their results for final classification. The main contributions of this paper are as follows:

- To improve the images and remove any flaws, the Top-hat and Bottom-hat filters are applied at the very beginning.
- A new layer called Entropic Field Propagation is proposed, replacing conventional pooling by leveraging spatial softmax, local entropy computation, and a learnable gating process. This layer preserves fine-grained spatial information and enhances feature learning in CNNs.
- An efficient Marine Predator Optimization technique has been proposed using the Newton-Raphson approach. Additionally, modifications are made to the Fitness and Activation functions.
- The best selected features are fused using a novel mean threshold-based fusion approach that ensures the fusion of unique features.

Literature review

Gastric cancer develops as a result of the abnormal stomach tissues growing larger. This type of tumor is the second most dangerous kind[22]. Cancer is a major public health concern as it is the second largest cause of death in human societies, after heart issues. Unhealthy diets, mineral sensitivity, smoking, and air pollution are all potential cancer-causing factors[23]. Researchers have developed several techniques for detecting and diagnosing stomach cancer, including Endoscopy and WCE[24]. Khan et al.[18]. presented a fully automated sequential method for classifying stomach diseases. To identify stomach infections, a variety of feature types, Including Feature Extraction, Fusion, and strong feature selection, were applied. In this article, three cases of stomach diseases and one healthy class are examined. The following datasets are utilized in this research to assess the suggested technique CVC-ClinicDB, ETIS-Larib, Kvasir-SEG, CUI Wah Private, and ASU-Mayo Clinic Colonoscopy Video Database. These datasets enabled the achievement of 99.46% accuracy. Using the transfer learning technique, the ResNet-101 pretrained model is enhanced for disease classification in gastroscopy images[25]. Based on the features of the images from the gastroscopy, these investigations classify stomach diseases and stomach cancer[26]. However, the CNN model still learned some redundant and irrelevant features that could limit its performance. Farah et al.[27]. presented a deep CNN-based method for feature extraction and optimization that can be used to classify gastroenteritis. The suggested method derived deep CNN features by transfer learning with DenseNet-201 and Inception V3. The proposed model classifies stomach disorders from medical images by utilizing feature fusion, feature optimization, data augmentation, and enhanced deep CNN models. Using a metaheuristic technique, redundant features are removed, and deep features are extracted using Inception v3 and DenseNet-201[28]. The optimized feature matrix, which has a 99.8% accuracy rate on the combined dataset for stomach diseases, was classified using machine learning techniques[29]. A comparison was conducted using modern approaches, and the findings show higher precision. The primary drawback of the implant method is the continuous increase in computing cost resulting from feature fusion. The development

of a large database and an advanced deep-learning model will be included in further work, particularly for the diagnosis of stomach diseases. J.V.Thomas et al.[30]. presented a method to identify and classify stomach issues by using transfer-learning models with a variety of pre-trained models. Compared to other modern methods, the suggested methods showed improvements in measures such as accuracy, precision, and recall. A publicly available multiclass Kvasir dataset was used throughout the entire experiment[31]. The best performance results were achieved by Efficient B0[32] with 98.01% accuracy, 98% precision, and 98% recall. The suggested approach can be applied to the detection and treatment of various diseases. Sanjeev et al.[33]. presented Several gastric-based diseases have been detected and classified using deep transfer learning models such as DenseNet201, EfficientNetB4, Xception, InceptionResNetV2, and ResNet152V2. Precision, loss, accuracy, F1 score, root mean square error, and recall were evaluated for these models. This study utilized the Kvasir dataset, and the Xception model achieved an accuracy of 98.2%. The primary issue with this study was the wide range of image sizes. Since most of the images had black boundaries, the performance of classification networks was constrained. Therefore, in the future, state-of-the-art image processing techniques can be utilized to enhance image quality, and an application can be developed to assist patients in quickly identifying the digestive issues they have. Nouman et al.[34]. presented an efficient contrast-enhancement methodology that optimizes brightness management to improve the contrast of WCE images, hence improving the classification of GI tract disorders. The proposed technique enhances the overall quality of WCE images by utilizing a genetic algorithm (GA) to adjust the fitness function and modify the contrast and brightness values within an image[35]. The proposed model achieved 96.40% accuracy, 93.02% recall, 97.57% precision, and 95.24% f-measure using the softmax classifier. Several fascinating aspects still require further research. In this study, for example, the impact of training multiple models or feature optimization was considered. These actions will be the focus of future studies, as they may improve performance. Javeria et al.[36]. presented a hybrid method that might accurately detect digestive system issues and encourage timely intervention, both of which would help to reduce the number of deaths. Dataset augmentation, preprocessing, feature extraction, fusion, optimization, and classification were the main stages of the suggested methodology. Transfer learning is used to extract Deep Learning features from the proposed XcepNet23 model and the ResNet18 model. This study uses two datasets, the hybrid dataset and the Kvasir V1 dataset. The hybrid dataset contains five classes, while the Kvasir V1 dataset has eight classes. The accuracy on the Kvasir V1 dataset was achieved at 99.24%.

Ahamed et al.[37]presented a deep learning architecture for the automated detection and segmentation of colorectal polyps using various combinations of YOLOv8 models for localization and a custom TR-SE-Net model for segmentation. TR-SE-Net was an architecture that utilized Squeeze-and-Excite (SE) blocks and a transformer-based encoder-decoder for efficient performance. The architecture was trained on the Kvasir-SEG dataset and was externally validated on CVC-ClinicDB, PolypGen, ETIS-LaribPolypDB, EDD 2020, and BKAI-IGH datasets. They found the best YOLOv8m model yielded 0.946 precision, 0.771 recall, 0.85 F1 score, 0.886 mAP50, and 0.695 mAP50 95, while TR-SE-Net yielded DSC of 0.8754, IoU of 0.7961, accuracy of 0.9647, and a superior speed of 54 FPS. They also developed a CAD system that combined both models to reduce the number of missed polyp detections. Finally, the authors reported severe problems with segmentation, generalizability, and complexity of deployment for real-time capabilities in a clinical instance, despite the high quality of their results. Ahamed et al.[38] presented a segmentation framework using an attention-guided MultiResUNet for polyp detection with endoscopic images. For the robustness of the model, the proposed technique employed Test Time Augmentation (TTA), utilizing horizontal and vertical flips to enhance generalization. The model was trained on a dataset (Kvasir-SEG) and evaluated on both datasets, Kvasir-SEG and CVC-ClinicDB, under TTA, attained an average of 0.8663 DSC, 0.8277 IoU, 0.9364 precision, 0.8060 recall, and 0.9993 accuracy. The model, taking up relatively low resources with 0.47 million parameters, is promising for a clinical setting. Weaknesses in this study are (1) the use of TTA provided performance increases, although not practicing actual performance, the augmentation process can become cumbersome, (2) the proposed technique practiced binary segmentation, thus evaluated on a binary polyp, further segmentation techniques and multiclass or generalization of polyp detection would be required in future research. Ahamed et al.[39] presented a lightweight classification framework that combined a Parallel Depthwise Separable CNN (PD-CNN) with Pearson Correlation Coefficient (PCC) for feature selection and an Ensemble Extreme Learning Machine (EELM) for classification of 27 diseases in the gastrointestinal (GI) tract. After being trained on the large-scale GastroVision dataset (8000 images), the proposed model had an accuracy of 87.75%, precision of 88.12%, recall of 87.75%, F1-score of 87.12%, and AUC-ROC of 98.89%. Additionally, the proposed model consisted of only 24 layers, with a total size of 9.79 MB and a testing time of 0.000001 s. The proposed model was efficient overall. The authors also implemented various explainability techniques to understand the decisions made by the proposed model. However, a limitation of this study is that it only focused on image classification, which does not include image segmentation or localization, both of which are important in disease diagnostics in the real world. Ahamed et al.[40] proposed a tri-stage deep learning architecture for GI diseases classification using a brand new Parallel Squeeze-and-Excitation CNN (PSE-CNN) as the feature extractor, Principal Component Analysis (PCA/TCA) for the parameters dimensionality reduction, and Deep Extreme Learning Machine (DELM) as the classification engine. Implemented and evaluation on the GastroVision dataset with a sample of 8000 images sourced from 27 disease classes, the architecture showed both an efficient architecture and effective performance with an overall accuracy of 97.24% in the first stage (3-class classification), 90.00% in the second stage (9-class) and 93.00% in the third stage (27-class). Their architecture is compact (14.88 MB) and has an inference time of 59.17 ms, which is clinically very relevant; however, due to the modeling complexity of multi-stage training and its use of PCA as a single static feature selector, it may not generalize well on unseen and changing datasets.

In summary, most researchers employed deep learning for the detection and classification of gastrointestinal (GI) and stomach conditions. A common challenge is feature redundancy and irrelevance when it is extracted from the CNN. Even with high accuracies, as with many other works, a reliance on fusion and optimization

techniques for better feature representations poses enormous computational time and overhead, making this problem very complex and ultimately much more challenging for real-time recommendation. Additionally, while many of the models included constraints based on data variability, several issues, such as inconsistent image resolution, resulted in black borders, which created additional constraints on the overall network performance. Lastly, many of the approaches have a complicated and inefficient use of static feature selection and processing techniques, such as PCA. Models constrained to static techniques become limited by dynamic or unseen data, resulting in constrained adaptations over time, thereby severely limiting their potential for robustness and scalability.

Proposed methodology

This section presents a proposed method for classifying stomach diseases using Marine Predator, a deep learning framework that applies modified convolutional neural networks with a new layer. The suggested methodology's architecture is displayed in Fig. 1. To improve contrast, we apply Top-hat and Bottom-hat filters to the original dataset images. Additionally, we employed three different methods of data augmentation, which involved operations such as rotation, flipping, and cropping, to increase the size and complexity of the dataset. After that, we modified the models (ResNet18 and ResNet50) and trained them on gastric intestinal disease images using transfer learning. We extract features from the deeper layer, using both of the optimized deep learning models. We applied the Marine Predator Algorithm, a proposed metaheuristic technique, to optimize both deep feature vectors extracted from the models. We then fuse the optimized feature vectors using a Threshold based fusion approach. Finally, Machine learning algorithms are applied to classify the final fused vector features for digestive tract diseases.

Collection of datasets

In this study, we investigate three datasets, Kvasir V1, Kvasir V2, and HyperKvasir, that contain images from inside the digestive tract[41]. The Kvasir V1 and V2 datasets comprise eight classes: normal pylorus, normal z-line, normal cecum, ulcerative colitis, polyps, esophagitis, and dyed and lifted polyps. The HyperKvasir dataset features 23 classes with 110,662 labeled images. This dataset is highly imbalanced, as the high-class category comprises 1,148 samples, while the lower-class category has only 6 samples. A brief description of the images in these datasets is provided in Table 1. We applied data augmentation techniques to perform the three methods shown in Fig. 2.

Contrast enhancement

Enhancing the input image is necessary to improve the quality of the original image. Despite recent advances in image capture technology, these devices still face numerous challenges. Some of these difficulties are specific to medical imaging, such as poor contrast, color distortions, and noise. Our suggested method aims to enhance the diagnosis of the affected region and improve the overall appearance of the image's pixel density. We consider combining the Top-hat and Bottom-hat filters because the low quality and contrast of the stomach cancer images necessitate this approach. Let $\sum^{\sim}$ represent a database containing a group of WCE photos. Consider $I(m,n)$ as a WCE image with $X \times Y \times K$ pixels, where $X = Y = 256\ and\ K = 3$, indicating the image's RGB format. The image's contrast is computed using the following Eqs. (1–4).

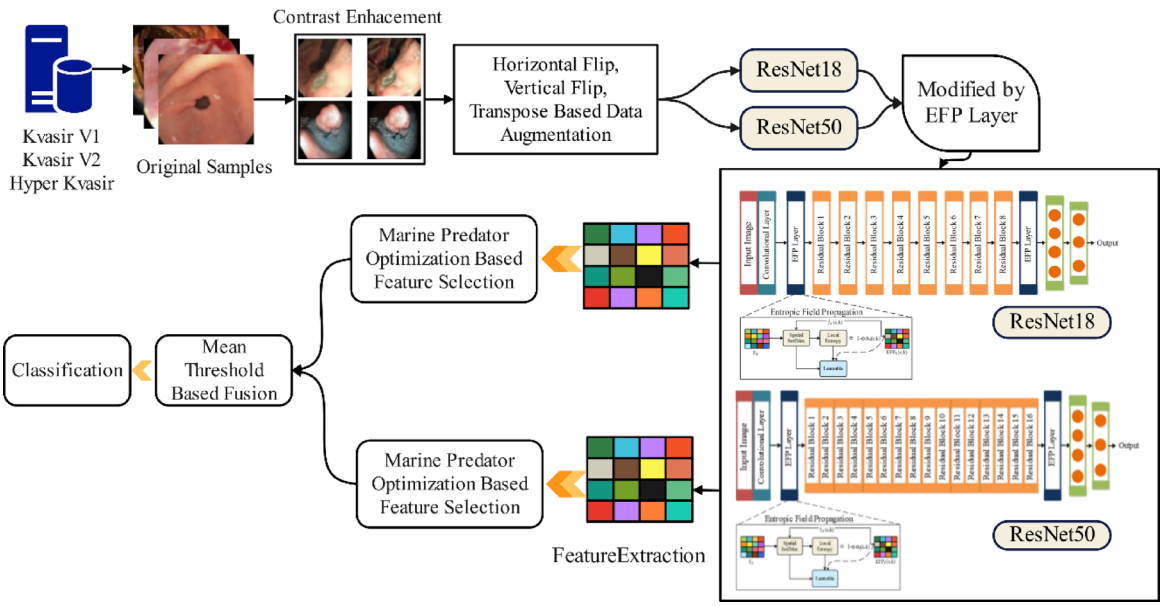

Fig. 1. The proposed architecture for classifying stomach diseases using deep learning and marine predator feature optimization.

Name of Dataset	Names of classes	Number of Images
Kvasir V1	Dyed and lifted polyps	500
	Dyed resection margins	500
	Polyps	500
	Esophagitis	500
	Normal pylorus	500
	Ulcerative-colitis	500
	Normal cecum	500
	Normal z-line	500
Kvasir V2	Dyed and lifted polyps	1000
	Dyed resection margins	1000
	Polyps	1000
	Esophagitis	1000
	Normal pylorus	1000
	Ulcerative-colitis	1000
	Normal cecum	1000
	Normal z-line	1000
HyperKvasir	bbps-2-3	1,148
	polyp	1,028
	cecum	1,009
	dyed-lifted-polyps	1,002
	pylorus	999
	dyed-resection-margins	989
	z-line	932
	retroflex-stomach	764
	bbps-0-1	646
	ulcerative-colitis-grade-2	443
	esophagitis-a	403
	retroflex-rectum	391
	esophagitis-b-d	260
	ulcerative-colitis-grade-1	201
	ulcerative-colitis-grade-3	133
	impacted-stool	131
	barretts-short-segment	53
	barretts	41
	ulcerative-colitis-grade-0-1	35
	ulcerative-colitis-grade-2-3	28
	ulcerative-colitis-grade-1-2	11
	ileum	9
	hemorrhoids	6

Table 1. Description of the collected datasets.

$$\Psi_{c1}(m,n) = Bth + Tth \tag{1}$$

$$Bth = Bth\,(\Psi\ (m,\ n) \cdot\ S) + Ct \tag{2}$$

$$Tth = Tth\,(\Psi\ (m,\ n) \circ\ S) +\ Ct \tag{3}$$

$$\phi\,\tilde{}_{c}(m,\ n) = (\Psi_{c1}(m,\ n)\ -\ \Psi\,(m,\ n))\ *\ Bth \tag{4}$$

Where, $\Psi_{c1}(m,n)$ is the resulting image's initial contrast improved, Bth is the bottom-hat converted values, Tth is the Top-hat converted values, $\phi\,\tilde{}_{c}(m,\ n)$ is the upgraded and enhanced contrast values, S is a structural element and Ct is a constant with a value of 1. Two more operators are used, such as closing and opening, use it and denote by $(\circ)\ and\ (\cdot)$. A few sample images of this step are shown in Fig. 3.

Dataset augmentation

To improve the performance of deep learning, several researchers employed data augmentation strategies[42]. Datasets for the deep learning model, which needs a lot of training data, are limited in the medical field[43]. To improve the diversity of the datasets, a data augmentation technique is needed. During the current study,

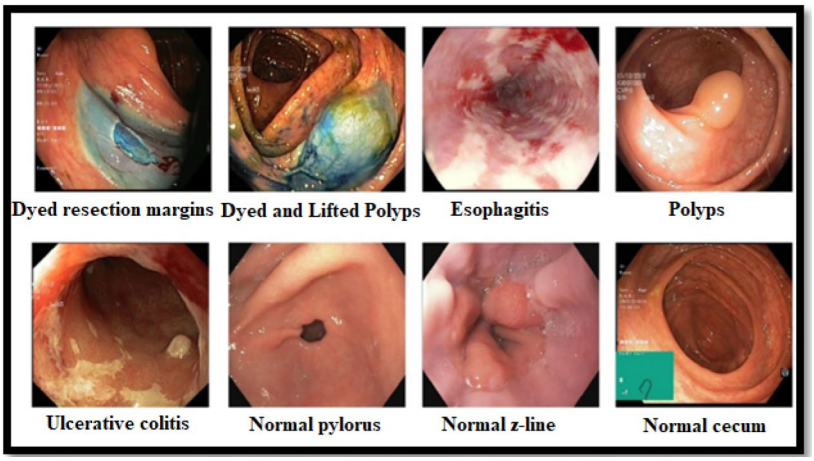

Fig. 2. A few sample images displaying distinct classes.

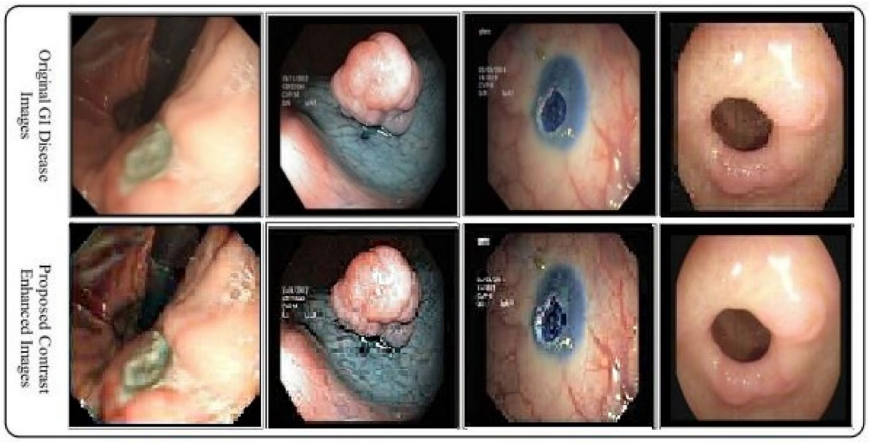

Fig. 3. Proposed contrast enhancing with WCE images.

two datasets were used for validation, as previously described. The first Kvasir V1 dataset consists of 4000 images that capture various gastrointestinal diseases. The dataset has eight classes of images, which are: normal-cecum, esophagitis, normal-z-line, normal-pylorus, normal-z-line, polyps, dyed-lifted-polyps, ulcerative-colitis, and dyed resection margins. Each class contains 500 images of different gastrointestinal conditions or regular appearances. The second dataset, Kvasir V2 of gastrointestinal images, contains 8000 images in eight classes. The classes are normal-ecum; esophagitis, normal-z-line, normal-pylorus, polyps, dyed-lifted-polyps, ulcerative-colitis, and dyed resection margins. They correspond to different gastrointestinal diseases or regular appearances. Two sets of the dataset were generated, with 70% allocated for training and 30% for testing. The original gastrointestinal images were trained for each class, but the dataset was insufficient for the deep learning model. To handle the limited samples and unbalance challenge, we augmented the dataset by performing three operations on the original images: horizontal flip, vertical flip, and rotation 90o. The mathematical formulation of these methods are: Suppose an input image is presented by $I_{(m,n)}$ with a dimension of $x \times y$ where $\{m \in (1,2,3,4,\ldots.,x)\}$ and $\{n \in (1,2,3,4,\ldots.,y)\}$. The three operations are performed on the input image to augment the data. Mathematically, this process is formulated using Eqs. (5–7).

$$I_{(m,n)}{}^{horF} = \psi_{m}(y+1-n) \tag{5}$$

$$I_{(m,n)}{}^{VerF} = \psi_{n}(x+1-m) \tag{6}$$

$$I_{(m,n)}{}^{Rot90} = \begin{bmatrix} cos90 & -sin90 \\ sin90 & cos90 \end{bmatrix} \begin{bmatrix} \psi_{m} \\ \psi_{n} \end{bmatrix} \tag{7}$$

Where $I_{(m,n)}{}^{horF}$, $I_{(m,n)}{}^{VerF}$, and $I_{(m,n)}{}^{Rot90}$ denoted the output of horizontal flip, vertical flip, and rotation 90-degree operations. This method enhances the variability of the dataset by introducing transformations that simulate different visual perspectives. This not only helps make the model more generalizable to unseen test

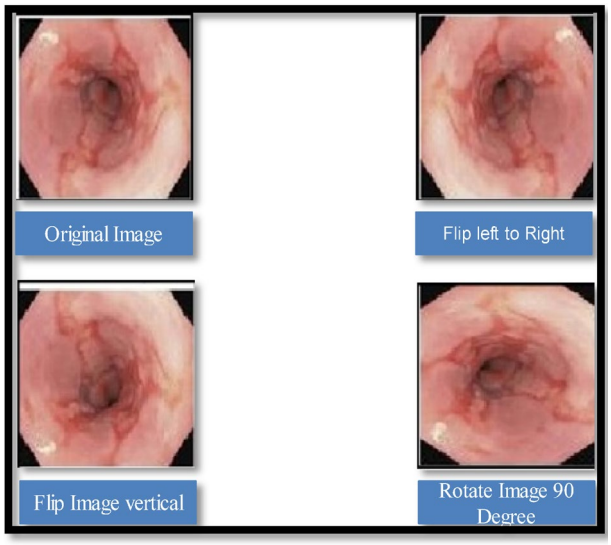

Fig. 4. visually on the selected dataset.

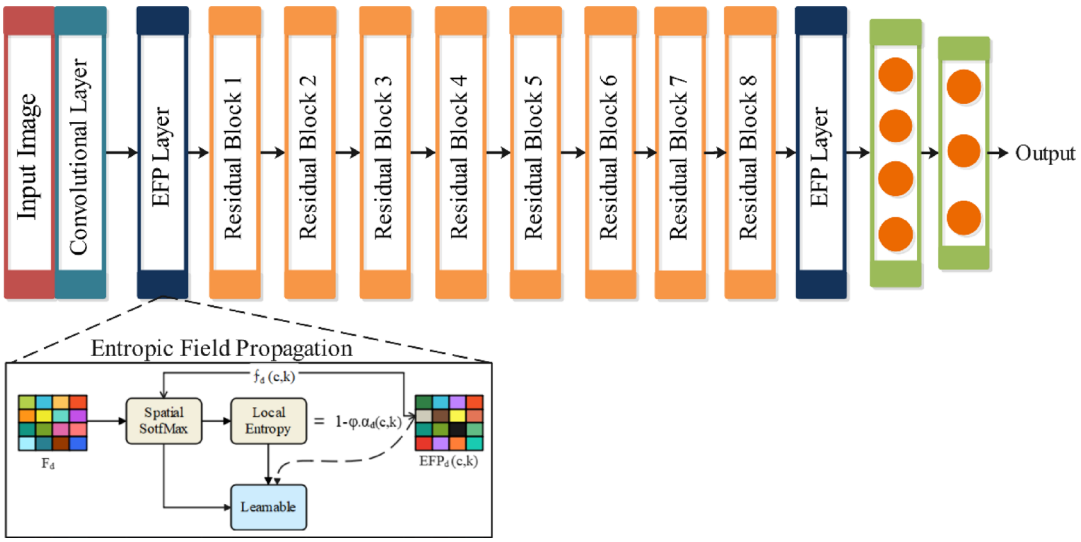

Fig. 5. Architecture of residual networks modified by the entropic field propagation (EFP) layer.

data, but also mitigates the risk of over-adaptation due to the limited size of the dataset. The visual representation of the augmentation process is shown in Fig. 4.

Modified ResNet (residual network) deep learning

ResNet18 and ResNet50, the two most widely implemented models for classifying medical images[44]. ResNet-18 has fewer layers than ResNet50, in total 18 layers; 17 convolutional layers and one fully connected layer. It performs a max pooling operation after performing a convolution layer with 7×7 Kernels and a stride of 2. There are four residual blocks in the architecture, each with increasing channel depths: 64, 128, 256, and 512. ResNet-50 is a CNN with 50 layers and a complex architecture. This model takes advantage of deeper hierarchical feature extraction through its layers, which starts with a 7×7 convolution and a 3×3 max pooling operation. In this study, we proposed a novel layer named Entropic Field Propagation (EFP) to modulate adaptively based on local entropy and prioritizing high-confidence (low-entropy) areas while suppressing the noise, meaning that local features propagate at different rates. The pooling layers are replaced with the EFP layers in both ResNet18 and ResNet50 networks, as shown in Fig. 5. This layer serves as both an information-theoretic filter, together with filtering for locality at every layer, while similarly drawing on the unique opportunities for spatial aggregation within pooling, the unpredictability of the spatial dimensions in the EFP layer is allowed to be determined by spatially adjacent pixels via attention. After replacing the pooling layer with EFP layers, both

modified ResNet18 and ResNet50 are trained on the selected datasets, and the deep features are extracted from the network using testing data.

Novelty 1: entropic field propagation (EFP) layer

The Entropic Field Propagation (EFP) Layer introduces a new mechanism in CNNs by recognizing that only some spatial areas in an image are relevant to the classification task. In Traditional CNNs, the pooling operations apply a fixed filter to the entire feature map, regardless of the quality of information in that local spatial area, as shown in Fig. 6. This uniform operation can lead to noisy or poor-quality features, especially in structurally complex images and low-quality medical imaging. Conversely, the EFP layer allows the model to modulate the propagation weights that feed-forward or back-propagate the propagated feature based on local entropy. The model and higher-level features would learn to down-weight areas of the image with high entropy and up-weight regions that are low entropy. Therefore, this dynamic behavior allows the network to focus on the most reliable features to improve robustness, interpretability, and classification accuracy. The mathematical formulation of this layer is:

Let $F \in \mathbb{R}^{H \times W \times D}$ is the input feature map, where H, W and D is the height. Width and channel depths, respectively. The EFP layer aims to adaptively filter features based on local entropy, which allows the model to highlight low-entropy regions and ignore uncertain areas. For the each channel $d \in \{1,2,3,4,\ldots,D\}$. The spatial features are normalized using the softmax activation over local neighborhood $\beta_{c,k}?F_d$, centered at each spatial location (c,k):

$$\oint_d (c,k) = \frac{e^{F_d(c,k)}}{\sum_{(m,n)?\beta_{c,k}} e^{F_d(m,n)}} \tag{8}$$

Where the $\oint_d (c,k)$ is the local distribution that approximates the importance of the features in the neighborhood. Next, to compute the uncertainty in the local neighborhood, Shannon entropy is measured at each spatial position.

$$\alpha_d (c,k) = -\sum_{(m,n)?\beta_{c,k}} \oint_d (m,n).\log\left(\oint (m,n)_d + \tau\right) \tag{9}$$

Where τ presented the constant values for the stability. This entropy captures the uncertainty of the feature responses around the reach location. Learnable parameters ϕ ? [0,1] is used to control the sensitivity of the model to entropy. The modulation weights are defined as:

$$w_d (c,k) = 1 - \phi .\alpha_d (c,k) \tag{10}$$

The weight attenuates the contribution of regions with entropy and increases the impact of low-entropy regions. In the final output of the EFP layer at the location (c,k) for depth d is measured by accumulating the neighboring features using the learnable filter $K_d \in \mathbb{R}^{z \times z}$, modulated by the softmax and entropy-based weights.

$$EFP_d (c,k) = \sum_{(m,n)\in\beta_{c,k}} K_d (m-c, n-k) . \oint_d (m,n).w_d (m,n) \tag{11}$$

Feature extraction

The Feature Extraction strategy may be helpful when we have a vast data collection and need to minimize waste without losing any crucial or valuable information. With the use of Feature Extraction, the number of duplicate data points in the data collection is reduced. In this study, the deep learning characteristics are obtained from the feature by deleting the final fully connected layer. Mathematically, these features are represented by $\mathbb{F}_1$ and $\mathbb{F}_2$.

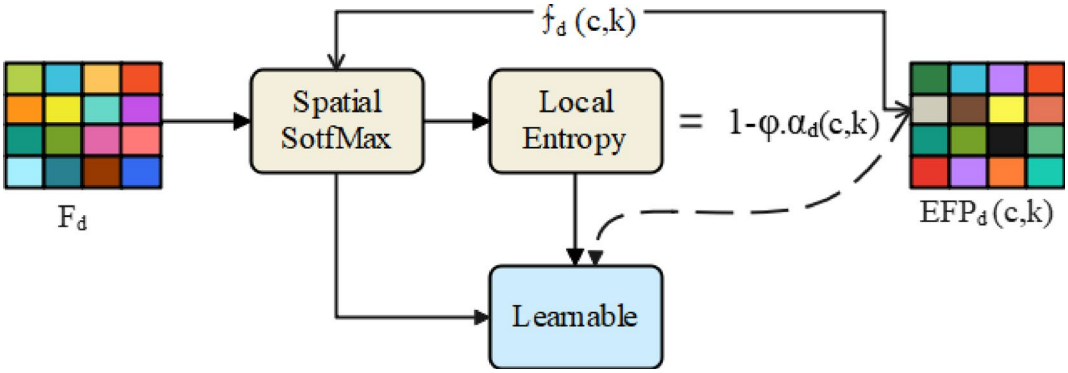

Fig. 6. Visualization of the entropic field propagation EFP layer.

The actual size of the extracted features is $N \times 2048$ and $N \times 512$, respectively. These vectors demonstrate that each image has 2048 features that were captured using modified ResNet50, and 512 features were extracted for each using ResNet18.

Marine predators algorithm

The motive behind MPA is seen in both predator-prey interaction and the overall hunting behavior of ocean Predators[45]. Here, a Predator aims at maximizing interaction rates to boost its chances of surviving in the wild. MPA uses two straightforward random walk algorithms to do a search: the Flight of Levy and Brownian motion. Meta-Heuristic Algorithms often use the Lévy flight. It works best when an active search is performed locally to prevent solution paralysis. Additionally, a well-known tool for global search is Brownian motion. Increasing the exchange level is necessary through means of extraction and investigation; developers combined the search efficiency of Lévy and Brownian motion.

Initialization: Using $Eq.\ (12)$ MPA chooses Nn search agents at random to begin the search.

$$xi^0 = lbi + q^{\cdot} * (ubi - lbi), i? \{1,2, \ldots ..\mathrm{Nn}\} \tag{12}$$

$q^{\cdot}$ is a random integer in the range $[0,1]$ where lbi and $ubi$are during initialization and the fundamental population matrix, lower and upper limits, the best-fitting search agents are included in a new $NnXDim$ matrix that is created. Population size and dimensions are indicated by Nn and Dim. The MPA calls it Elite.

$$\mathrm{Elite} = \begin{bmatrix} X1^l, 1 & X1^l, 2 \ldots\ldots\ldots X1^l & 1, \mathrm{Dim} \\ X2^l, 1 & X2^l, 2 \ldots\ldots..X2^l & 2, \mathrm{Dim} \\ XNn^l, 1 & XNn^l, 2 \ldots\ldots..XNn^l & 3, \mathrm{Dim} \end{bmatrix} \tag{13}$$

Where X^I indicates the highest Fitness Vector. Prey and Elite are comparable. Predators use it to update their positions. The beginning produces the initial Prey, from which the strongest person quickly becomes the Elite. The next paragraph explains how The Prey is shown.

$$\mathrm{Prey} = \begin{bmatrix} X1, 1 & X1, 2 \ldots\ldots\ldots X1 & 1, \mathrm{Dim} \\ X2, 1 & X2, 2 \ldots\ldots..X2 & 2, \mathrm{Dim} \\ XNn, 1 & Nn, 2 \ldots\ldots..XNn & 3, \mathrm{Dim} \end{bmatrix} \tag{14}$$

Zm, n Indicates the nth dimension of the m predator. The optimization procedure depends on these two matrices. The primary iterative search method starts after startup. It is divided into three phases, each representing different Predator-Prey situations and employing various search strategies. Iterations form the basis of the three stages. $it\ ?\ itmax$ and $itmax$ where $itmax$ stands for the maximum number of iterations. During this phase, MPA updates possible solutions.

Stage 1

High speed $(it < itmax3it < itmax3)$ represents a scenario in which the Predator is being hurried by the Prey. This strategy builds upon the study and utilizes more of the earlier phases. $Eq.\ (15)$ carries out the mathematical expression of this rule.

$$Stepsize\ j = \boldsymbol{R}r\ x\ (Elite\ j - \boldsymbol{R}r\ \ x\ predator\ n \tag{15}$$

$$Predator\ n = \ Predator\ n\ + S.\boldsymbol{T}\ x\ Stepsize\ n \tag{16}$$

Where $[0,\ 1]$ is a range for the random number Rr, The normal distribution is used to determine Brownian motion. R is a vector of [0, 1] uniformly distributed integers. While the constant P has a value of [0.5]. During this phase, Predators and Prey move very quickly, making it easier to explore remote portions of the searching location.

Stage 2

$(1/3\ itmax,\ it,\ and\ 2/3\ itmax)$ Both the Prey and the Predator move quickly. In this phase, the Transmission step shifts from study to utilization. As a result, the population is divided into two groups: Predators that take advantage of Brownian motion and Prey that are explored using Lévy motion. $Eq.\ (17)$ was employed to determine who the first half of the population was in $Eq.\ (18)$.

$$Stepsize\ n = \boldsymbol{R}levy\ x\ \ (Elite\ n - \boldsymbol{R}levy\ \ x\ prey\ n),\ n? \{Nn\} \tag{17}$$

$$Prey\ n = prey\ n + S.\boldsymbol{T}\ x\ Stepsize\ n \tag{18}$$

Using the Lévy assumption, where Rlevy is a random number. While increasing the step size to the Prey location influences prey movement, multiplying Rlevy, Prey n imitates the predator motion in Lévy.

$$Stepsize\ j = \ \boldsymbol{R}r\ x\ \ (\boldsymbol{R}r\ x\ Elite\ n\ - prey\ n),\ n? \{\ Nn\} \tag{19}$$

$$Prey\ n = \ Elite\ n\ + P * KL\ x\ Stepsize\ n \tag{20}$$

Where KL is a $step\ size$ control parameter that may be computed as follows:

$$KL = (1 - it/itmax)^{2.t/max} \tag{21}$$

Stage 3
Slow speed $(it > 2/3itmax)$: The Lévy flight can alter the population :

$$Stepsize\ n = \boldsymbol{R}levy\ x\ (\boldsymbol{R}\ levy\ x\ Elite\ n\ - prey\ n),\ n?\ \{Nn\} \tag{22}$$

$$Prey\ n = \ Elite\ n\ + P * KL\ x\ Stepsize\ n \tag{23}$$

The concept of marine predators, known as flow creation or Fish Aggregating Devices (FADs), is utilized to provide variety to potential solutions using the formulation below. These predators make longer jumps to different locations in search of more food.

$$Prey\ n = \ \{Prey\ n + \ K[lbm + rx(ubm - lbm)xl \tag{24}$$

$$Prey\ n + [FADs(\ 1 - r) + (preyr1 - preyr2\ )\} \tag{25}$$

Where

$$prey \rightarrow nprey \rightarrow n,\ prey \rightarrow -r1prey \rightarrow r1, \tag{26}$$

and

$$prey \rightarrow \ r2prey \rightarrow \ r2 \tag{27}$$

Represent vectors for the nth candidate solution, a random finding solution, and a second random finding solution, in that order; where [0,1] is the range for the random number r, FADs is a constant equal to 0.2, and ll denotes a binary vector that contains both zero and one. The modified marine predator is employed on both extracted feature vectors to select the best features. The deminsion of optimized features are $N \times\ 1147$ and $N \times\ 384$.

Novelty 2: mean threshold based fusion

We proposed a novel fusion approach named Mean threshold-based fusion, which fused information of selected feature vectors. The length of vectors is computed as:

$$\mathbb{F}_{length} = \sum\ (\mathbb{F}_1\ (length)\ , \mathbb{F}_2\ (length)) \tag{28}$$

There is a significant chance that unnecessary and duplicate features occur due to the considerable feature length. Using a mean threshold function, we attempted to reduce this problem. For the next step, we compared the mean value of each feature in this function. It is defined mathematically as follows:

$$\mathbb{F}_{fu} = \begin{cases} \mathbb{F}_{fu}\ (k) & if\ \mathbb{F}_{length}\ (i) \geq\ m \\ Ignore & Elsewhere \end{cases} \tag{29}$$

According to this function, the features in the fused vector that are equal to or larger than m are selected, and the process proceeds to the next stage. The remaining features are ignored. The best features are then chosen using the Marine Predator Algorithm. The following is a complete explanation of the MPA.

Results and discussion

We examine the efficiency of our suggested framework using two separate datasets of WCE images that include different GI diseases. We use two datasets of WCE images, namely Kvasir V1 and Kvasir V2, for our experiments. We split each dataset into two parts: a test set comprising 30% of the images and a training set containing 70% of the images. We use the training set to train our models and the test set to evaluate their performance. In this study, we perform 10-fold cross-validation on each of our experiments. We use the following hyperparameters to train our DL models: a small batch size of 60, 100 epochs, an SGD optimizer with a learning rate of 0.0001 and a momentum of 0.75. We extract features from images using the sigmoid function, and we test the classification accuracy using the cross-entropy loss function. We employ ten classifiers, comprising three SVMs, two KNNs, and five NNs, to evaluate the accuracy and time efficiency of our framework. On a computer with a 32 GB graphics card and 8 GB of RAM, we execute our framework using MATLAB R2024a as the simulation platform. To predict intestinal diseases, the optimal algorithm was chosen using machine learning techniques based on its accuracy and speed. We evaluated the efficiency of the method using metrics including accuracy, precision, recall, F1-score, FNR, and time.

We use feature extraction to select the original feature, and we evaluate its performance using five metrics: precision, sensitivity, accuracy, FNR, and time. The outcomes of experiment 1 employing the ResNet 18 model are shown in Tables 2 and 3, respectively, for the Kvasir datasets V1 and V2. The accuracy of the cubic SVM classifier is 97.3%, whereas that of the quadratic SVM classifier is 97.1%. The accuracy of the competing classifiers is lower. The confusion matrix for the cubic SVM classifier, which can differentiate eight classes, as shown in Fig. 7, includes normal z-line, normal pylorus, dyed resection margins, normal cecum, dyed-lifted

Classifier	Precision	Sensitivity	F1-Score	FNR	Accuracy	Time(s)
CSVM	**97.32**	**97.3**	**97.3**	**2.7**	**97.3**	**91.744**
QSVM	97.1	97.06	97.0	2.938	97.1	66.853
MGSVM	97.02	97.02	97.0	2.98	97.0	112.84
LSVM	96.78	96.77	96.7	3.23	96.8	55.071
FKNN	96.66	96.66	96.6	3.34	96.7	54.753
CGSVM	97.01	96.72	96.8	3.28	96.7	86.533
WNN	96.61	96.61	96.6	3.39	96.6	34.506
MNN	96.37	96.37	96.3	3.63	96.4	21.245
WKNN	96.20	96.12	96.1	3.88	96.1	54.252
BNN	95.91	95.93	95.9	4.07	95.9	34.648

Table 2. Kvasir V1 classification results using the proposed feature ResNet-18. Bold values denote the most significant value.

Classifier	Precision	Sensitivity	F1-Score	FNR	Accuracy	Time(s)
CSVM	**93.07**	**88.16**	**90.54**	**11.84**	**88.2**	**90.41**
QSVM	88.77	88.0	88.38	12	88.0	55.29
MGSVM	87.9	87.13	87.51	12.87	87.2	111.63
LSVM	86.35	85.98	86.16	14.02	86.0	82.03
FKNN	86.08	85.75	85.91	14.25	85.7	81.06
CGSVM	85.6	85.43	85.51	14.57	85.4	690.25
WNN	85.22	84.98	85.10	15.02	85.0	379.37
MNN	85.15	84.97	85.05	15.03	85.0	407.8
WKNN	84.63	83.0	83.80	17.0	83.0	56.01
BNN	83.62	82.43	83.02	17.57	82.4	53.32

Table 3. Kvasir V2 classification results using proposed feature ResNet-18. Bold values denote the most significant value.

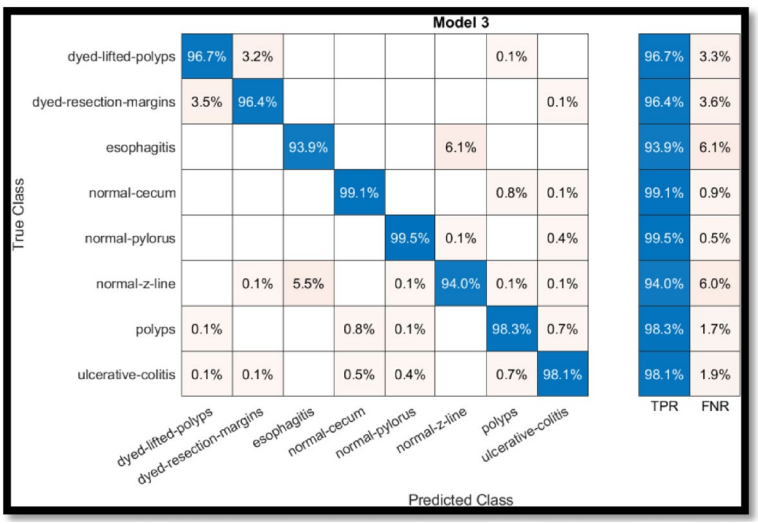

Fig. 7. Cubic SVM classifier's confusion matrix.

polyps, and esophagitis. Figure 8 compares the suggested features on the Kvasir V1 and Kvasir V2 datasets in terms of computational time.

The outcomes of Experiment 2, employing the ResNet-50 model, are presented in Tables 4 and 5 for the Kvasir V1 and V2 datasets, respectively. The coarse Gaussian SVM classifier comes in second with 96.4% accuracy, while the cubic SVM classifier achieves the highest accuracy at 96.5%. Other classifiers are less accurate. The confusion matrix for the cubic SVM classifier, which can differentiate between eight classes, is shown in Fig. 9, including

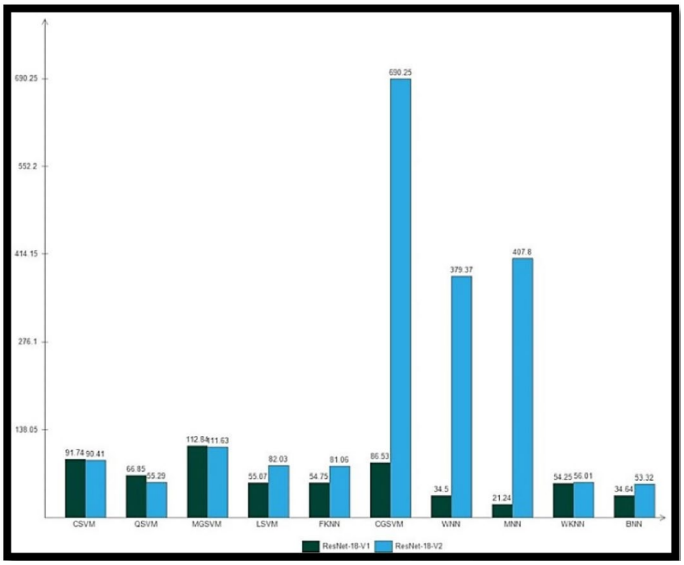

Fig. 8. Computational time-based comparison among different classifiers.

Classifier	Precision	Sensitivity	F1-Score	FNR	Accuracy	Time(s)
CSVM	**96.53**	**96.52**	**96.5**	**3.48**	**96.5**	**389.41**
QSVM	96.4	96.36	96.3	3.64	96.4	330
WNN	95.67	96.4	96.0	3.6	95.7	112.45
MGSVM	95.51	95.46	95.4	4.54	95.5	648.08
LSVM	95.47	95.36	95.4	4.64	95.4	217.52
MNN	95.17	95.16	95.1	4.84	95.2	73.478
FKNN	94.71	94.65	94.6	5.35	94.7	215.87
BNN	94.64	94.64	94.6	5.36	94.6	126.29
WKNN	94.28	93.97	94.1	6.03	94.0	210.64
CGSVM	94.25	93.97	94.1	6.03	94.0	474.79

Table 4. Kvasir V1 classification results using proposed feature ResNet-50. Bold values denote the most significant value.

Classifier	Precision	Sensitivity	F1-Score	FNR	Accuracy	Time(s)
CGSVM	**91.41**	**86.6**	**88.9**	**13.4**	**86.6**	**480.11**
LSVM	86.78	85.9	86.3	14.1	85.9	216.33
QSVM	85.08	84.53	84.8	15.47	84.5	350.42
C SVM	84.9	84.42	84.6	15.58	84.4	402.2
MGSVM	85.075	84.43	84.7	15.57	84.4	586.06
MNN	84.28	84.05	84.1	15.95	84.0	1380.2
WNN	84.25	84.025	84.1	15.97	84.0	564.35
BNN	83.47	83.31	83.3	16.69	83.3	126.6
WKNN	82.73	81.1	81.9	18.9	81.1	206.54
FKNN	80.68	79.48	80.0	20.52	79.5	206.83

Table 5. Kvaris V2 classification results using proposed feature ResNet-50. Bold values denote the most significant value.

normal z-line, normal pylorus, normal cecum, dyed-lifted polyps, dyed resection margins, and esophagitis. Figure 10 compares the suggested features on the Kvasir V1 and V2 datasets in terms of computational time.

In experiments 3 and 4, using ResNet 18- Kvasir V1 and V2 datasets, we applied our Marine Predator Algorithm on a feature vector and evaluated different classifiers. We also performed serial-based fusion after

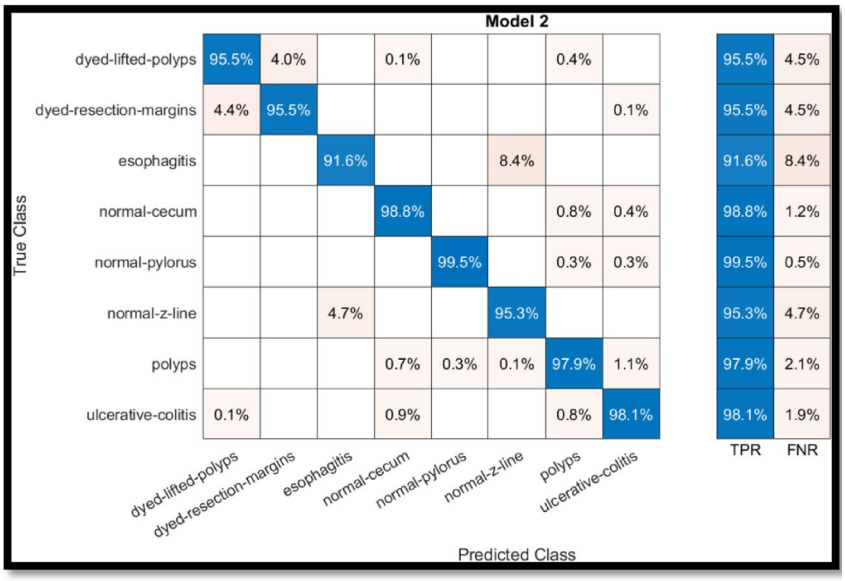

Fig. 9. Cubic SVM classifier's confusion matrix.

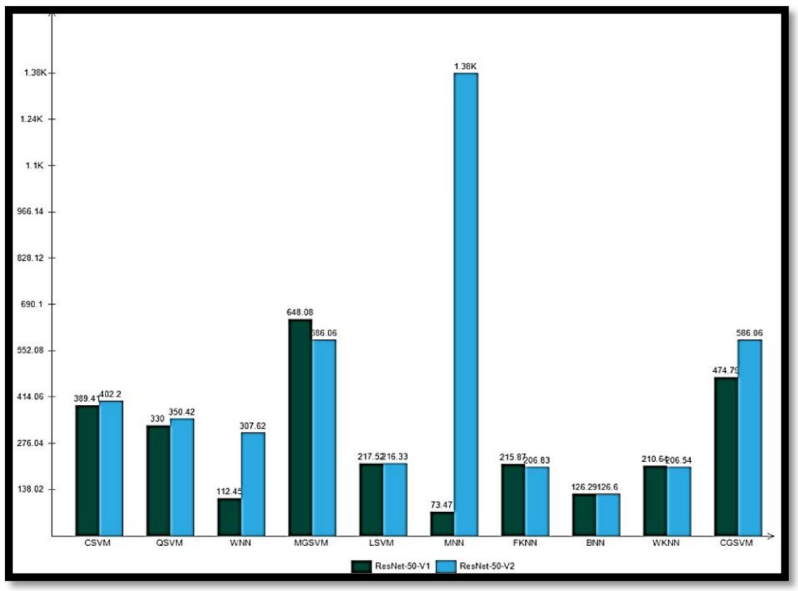

Fig. 10. Computational time-based comparison among different classifiers for Resnet50 on selected classifiers.

Marine Predator Optimization to enhance the results. Tables 6, 7 and 8 provides the summary of each classifier's accuracy, sensitivity, precision, and time. The best accuracy before fusion was 97.2% by Cubic SVM, which took only 29.93 s. The best accuracy after fusion was 99% by Linear SVM. The fusion improved the accuracy and processing time of all classifiers. Comparison of the suggested fusion and selection processes for the Kvasir V1 and V2 datasets in terms of computational time. We get to the conclusion that our serial-based fusion following Marine Predator Optimization is efficient in terms of accuracy and speed (can be seen in Fig. 11).

In experiments 5 and 6 Using ResNet 50- Kvasir V1 and V2 datasets, we applied our Marine Predator Algorithm on a feature vector and evaluated different classifiers. We also performed Serial based Fusion after Marine Predator Optimization to enhance the results. Tables 9, 10 and 11 provide a summary of each classifier's accuracy, sensitivity, precision, and time. The best accuracy before fusion was 96.2% by Cubic SVM on ResNet 50 V1 and 86.1 on ResNet 50 V2, which took only 139.03 s. The best accuracy after fusion was 89.6% by coarse Guassian SVM. The fusion increased the accuracy and decreased the time for all classifiers. On ResNet 50, a comparison of the proposed fusion and selection processes' computing times for the Kvasir dataset's versions V1 and V2. We get to the conclusion that our serial based fusion following Marine Predator Optimisation is efficient in terms of accuracy and speed (can be seen in Fig. 12).

Classifier	Precision	Sensitivity	F1-Score	FNR	Accuracy	Time(s)
CSVM	**97.24**	**97.17**	**97.2**	**2.83**	**97.2**	**29.932**
MGSVM	97.1	97.1	97.1	2.9	97.1	36.163
QSVM	96.98	97.0	96.9	3	97.0	27.256
LSVM	96.58	96.58	96.5	3.42	96.6	24.784
CGSVM	96.38	96.35	96.3	3.65	96.4	34.456
WKNN	96.02	95.98	96.0	4.02	96.0	19.088
WNN	95.91	95.92	95.9	4.08	95.9	18.707
MNN	95.87	95.85	95.8	4.15	95.9	17.997
FKNN	95.77	95.76	95.7	4.24	95.8	22.281
BNN	94.9	94.91	94.9	5.09	94.9	27.231

Table 6. The proposed algorithm for selecting the best features ResNet-18, dateset V1. Bold values denote the most significant value.

Classifier	Precision	Sensitivity	F1-Score	FNR	Accuracy	Time(s)
CGSVM	93.07	88.32	90.6	11.68	88.3	42.155
LSVM	88.78	87.87	88.3	12.13	87.9	27.034
MSVM	87.32	86.62	86.9	13.38	86.6	51.616
QSVM	86.30	85.91	86.1	14.09	85.9	35.98
CSVM	85.72	85.28	85.4	14.72	85.3	40.444
WNN	84.71	84.52	84.6	15.48	84.5	408.19
BNN	84.55	84.38	84.4	15.62	84.4	240.62
MNN	84.20	84.0	84.0	16	84.0	268.53
WKNN	84.52	83.08	83.7	16.92	83.1	25.762
FKNN	82.10	81.05	81.5	18.95	81.0	26.245

Table 7. The proposed algorithm for selecting the best features ResNet-18 dataset V2.

Classifier	Precision	Sensitivity	F1-Score	FNR	Accuracy	Time(s)
LSVM	98.95	98.96	98.9	1.04	99.0	97.48
QSVM	98.78	98.9	98.8	1.14	98.9	135.63
CSVM	98.93	98.95	98.9	1.05	98.9	155.83
MGSVM	98.77	98.76	98.7	1.24	98.9	254.17
WNN	98.66	98.65	98.6	1.35	98.7	46.54
MNN	98.33	98.35	98.3	1.65	98.4	28.642
CGSVM	98.42	98.35	98.3	1.65	98.4	178.62
BNN	98.3	98.31	98.3	1.69	98.3	45.235
WKNN	97.0	96.78	96.8	3.22	96.8	98.79
FKNN	96.81	96.76	96.7	3.24	96.8	100.19

Table 8. Classification results of the fusion-based classification.

Model interpretability

In this section, the interpretability of the proposed model has been evaluated using Grad-CAM, as shown in Fig. 13. Grad-CAM aims to visualize and understand the decision process of the modified ResNet50 model. Each row consists of the original endoscopic images and their respective Grad-CAM visualizations, demonstrating how the model focuses on specific areas in the GI image. Each red circle in the original images indicates clinically significant areas, such as polyps, lesions, or inflammation, that are relevant for the model's diagnosis. In the Grad-CAM-generated images, the thermal color scheme corresponds to the model's scores for specific areas of the image. Regions in red and yellow indicate high activation, green indicates moderate activation, and blue indicates that the area had little importance. The visualizations indicate that the model focuses on appropriate pathological areas of the images, meaning it can learn relevant features from the image and make more reliable predictions. Interpreting how the model reaches its decisions provides transparency, which will increase confidence in its clinical usefulness and in the decision to use EFP to improve accuracy and interpretability.

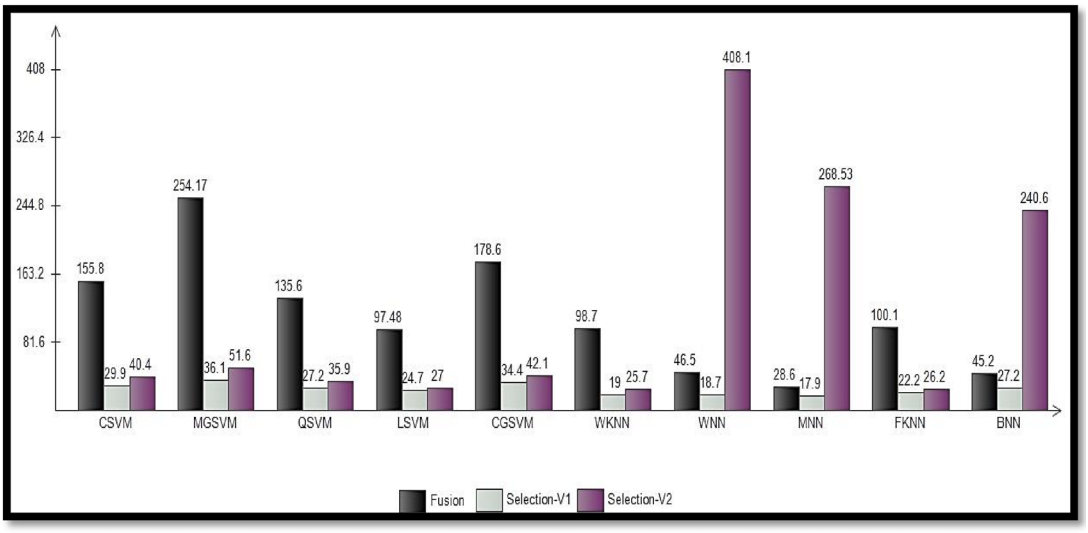

Fig. 11. Comparison of proposed fusion and selection steps in terms of computational time of ResNet-18 for Kvasir V1 and V2 datasets.

Classifier	Precision	Sensitivity	F1-Score	FNR	Accuracy	Time(s)
CSVM	**96.25**	**96.25**	**96.2**	**3.75**	**96.2**	**140.48**
QSVM	96	95.92	95.9	4.08	95.9	126.73
MGSVM	95.55	95.48	95.5	4.52	95.5	209.3
LSVM	95.4	95.3	95.3	4.7	95.3	87.071
WNN	95.11	95.1	95.1	4.9	95.1	55.698
MNN	94.51	94.5	94.5	5.5	94.5	36.52
FKNN	94.46	94.42	94.4	5.58	94.4	85.311
CGSVM	94.42	94.16	94.2	5.84	94.2	165.7
WKNN	94.17	93.91	94.0	6.09	93.9	85.009
BNN	93.62	93.61	93.6	6.39	93.6	96.445

Table 9. The proposed algorithm for selecting the best features ResNet-50 dateset V1. Bold values denote the most significant value.

Classifier	Precision	Sensitivity	F1-Score	FNR	Accuracy	Time(s)
CGSVM	**82.62**	**86.11**	**84.3**	**13.89**	**86.1**	**130.92**
LSVM	86.7	85.8	86.2	14.2	85.8	70.341
MGSVM	85.07	84.43	84.7	15.57	84.5	158.48
QSVM	84.45	83.86	84.1	16.14	83.9	103.21
CSVM	84.21	83.56	83.8	16.44	83.5	114.51
MNN	83.03	82.65	82.8	17.35	82.7	82,667
WNN	82.95	82.55	82.7	17.45	82.5	564.35
BNN	82.62	82.4	82.5	17.6	82.4	436.75
WKNN	82.52	81.43	81.9	18.57	81.9	66.548
FKNN	79.92	78.9	79.4	21.1	78.9	66.562

Table 10. The proposed algorithm for selecting the best features ResNet-50 dateset v2. Bold values denote the most significant value.

Ablation study

The ablation study shown in Fig. 14 details the performance effects of the proposed fusion methods on classification accuracy using the Kvasir v1 and v2 data with the specified classifier. First, it is clear that Mean Threshold Fusion consistently outperforms Traditional Fusion for both datasets. When comparing the classifiers, we observe that SVM models (LSVM and CGSVM) are the only approaches that achieve higher accuracies and

Classifier	Precision	Sensitivity	F1-Score	FNR	Accuracy	Time(s)
CGSVM	94.02	89.57	91.7	10.43	**89.6**	139.03
LSVM	89.91	89.13	89.5	10.87	89.1	81.972
MSVM	89.22	88.48	88.8	11.52	88.5	180.34
MNN	87.41	87.35	87.3	12.65	87.3	142.04
BNN	87.26	87.17	87.2	12.83	87.2	435.89
QSVM	87.23	86.96	87.0	13.04	87.0	117.72
WNN	86.95	86.87	86.9	13.13	86.9	158.24
CSVM	86.71	86.4	86.5	13.6	86.4	130.44
WKNN	86.8	85.61	86.2	14.39	85.6	79.203
FKNN	83.82	82.56	83.1	17.44	82.5	80.973

Table 11. Classification results of the fusion-based classification. Bold values denote the most significant value.

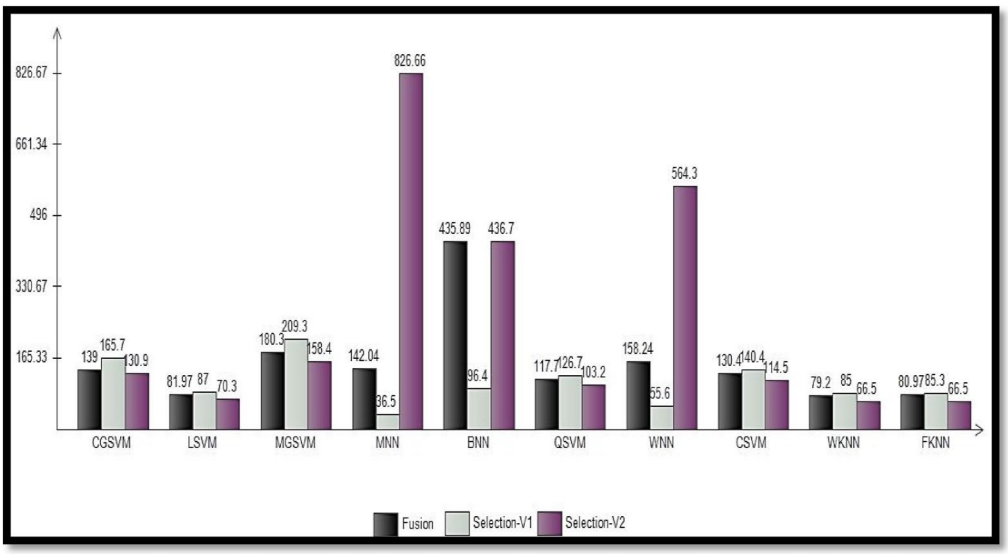

Fig. 12. Comparison of proposed fusion and selection steps in terms of computational time of ResNet-50 for Kvasir V1 and V2 datasets.

generalize well to all data. In contrast, WNN, MNN, and BNN models achieve high accuracy but exhibit lower generalizability compared to SVM models. The worst classifiers are WKNN and FKNN, which consistently yield the lowest accuracy, indicating that the K-NN approach is not well-suited to this type of classification task. It is also important to note that the overall accuracy of classification is higher on Kvasir v1 than on Kvasir v2, which could suggest that the Kvasir v2 data has more variability (noise) incorporated into the classes. Therefore, combining the SVM classifiers with the Mean Threshold Fusion method provides the most reliable and effective strategy for robust classification in real-world contexts.

Model generalizability

In this experiment, the model generalizability has been evaluated. The hyperKvasir dataset is employed, and classification results are shown in Table 12. According to the table, across all classifiers, it is apparent that performance improves when using features from Modified ResNet50 compared to Modified ResNet18. The deeper and more expressive features of ResNet50 have helped improve classification. For instance, the Gaussian CGSVM classifier yields an increase in F1-score from 71.7 to 81.74 and accuracy from 77.6 to 84.51% when changing from ResNet18 features to ResNet50 features. Similar trends are observed in the other classifiers, such as LSVM, MSVM, and MNN, where each achieves enhanced precision, sensitivity, and accuracy with ResNet50 compared to ResNet18. Interestingly, classifiers such as the MNN and WKNN (with ResNet50) achieved consistent F1 scores and metrics indicating balanced performance and robustness to both class imbalance and feature variability. The MNN achieved an overall precision of 83.14%, sensitivity to 82.14% and accuracy of 84.39% indicating strong consistency. The FKNN achieved the highest sensitivity amongst all classifiers with ResNet50 features (83.41%), which is significant for medical applications. Notably, although the decline in features improved the performance of most classifiers, WNN showed minimal improvement, with only a modest increase in accuracy (73.1–79.82%) and a slight detriment in the F1-score. The WNN classifier may struggle to utilize high-dimensional features and remain sensitive to noise.

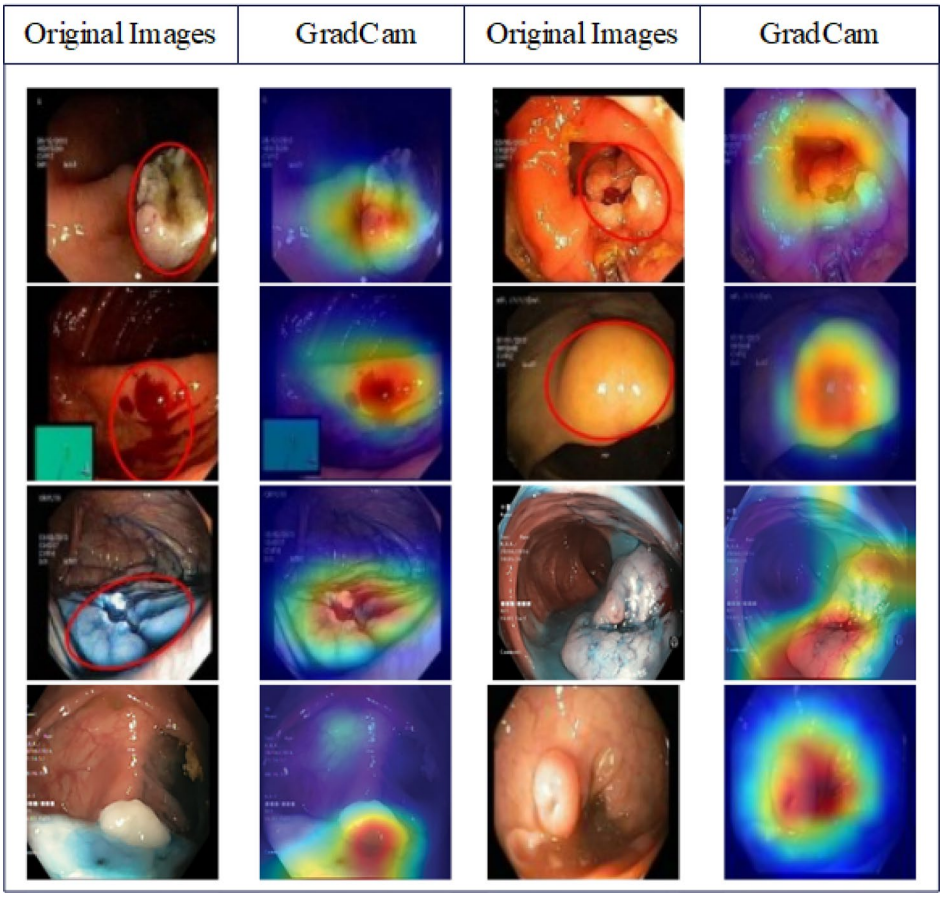

Fig. 13. Model interpretation using the GradCam visualization.

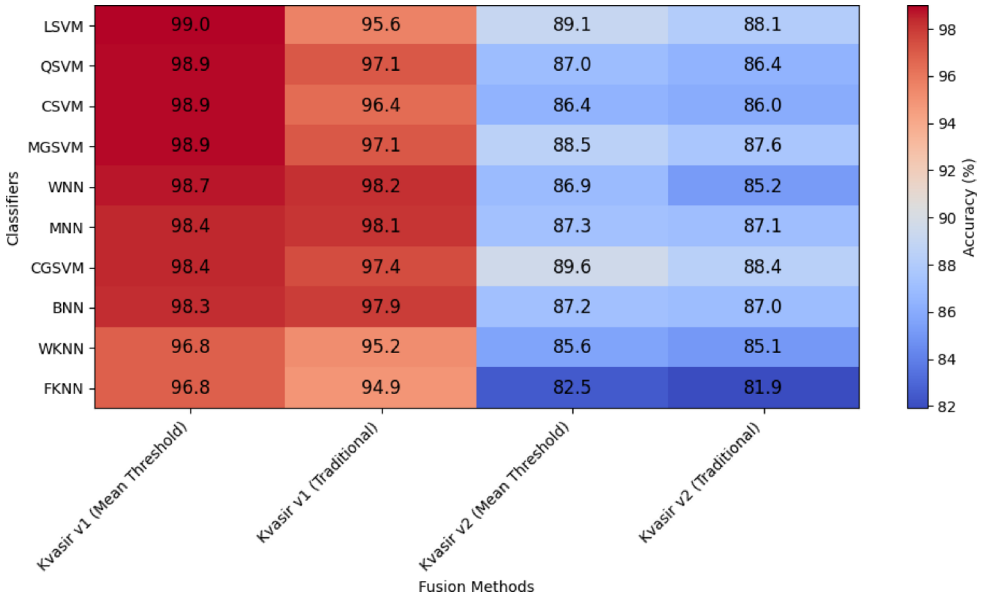

Fig. 14. Ablation study based on the novel mean threshold-based fusion vs. traditional fusion (serial Fusion).

Classifier	Modified ResNet18	Modified ResNet50	Precision	Sensitivity	F1-Score	Accuracy
CGSVM	✓		74.02	72.57	71.7	**77.6**
		✓	84.14	82.41	81.74	**84.51**
LSVM	✓		73.91	71.13	70.5	75.5
		✓	83.19	81.17	80.14	83.12
MSVM	✓		75.22	71.48	73.8	76.3
		✓	83.23	81.13	83.65	82.24
MNN	✓		73.41	75.35	72.3	73.1
		✓	83.14	82.14	82.89	84.39
BNN	✓		71.26	74.17	71.2	71.2
		✓	80.62	80.17	80.14	80.77
QSVM	✓		72.23	74.96	71.0	72.4
		✓	79.37	80.74	78.74	80.81
WNN	✓		73.95	73.87	73.9	73.1
		✓	75.94	72.56	71.11	79.82
CSVM	✓		71.71	71.4	71.5	71.4
		✓	81.92	79.78	79.77	81.97
WKNN	✓		74.8	75.61	72.2	75.4
		✓	81.71	80.31	81.79	82.71
FKNN	✓		71.82	70.56	70.1	70.6
		✓	81.23	83.41	80.63	83.67

Table 12. Performance of models' generalizability using Hyperkvasir dataset. Bold values denote the most significant value.

model variant	Training accuracy (%)	Training time (hrs)
Kvasir 1		
ResNet18	95.54	1 h 27 min
ResNet50	95.11	1 h 51 min
ResNet18 + EFP Layer	97.00	1 h 6 min
ResNet50 + EFP Layer	96.54	1 h 37 min
Kvasir 2		
ResNet18	84.02	1 h 49 min
ResNet50	85.81	2 h 3 min
ResNet18 + EFP Layer	88.34	1 h 21 min
ResNet50 + EFP Layer	86.65	1 h 43 min
HyperKvasir		
ResNet18	73.51	3 h 14 min
ResNet50	81.97	3 h 47 min
ResNet18 + EFP Layer	77.41	2 h 41 min
ResNet50 + EFP Layer	84.57	3 h 2 min

Table 13. Proposed model evaluations using before and after EFP layer modification.

Impact of EFP layer

In this experiment, the impact of the EFP layer is evaluated, as shown in Table 13. From the Kvasir 1 dataset, we can observe that the training accuracy of both ResNet18 and ResNet50 has improved considerably. ResNet18, combined with an EFP layer, improved training accuracy from 95.54 to 97.00%, achieving better performance while reducing training time. The results were similar for the Kvasir 2 dataset; ResNet18 improved from 84.02 to 88.34% accuracy, with a shorter training time. There were incremental results for ResNet50 that had incorporated EFP. HyperKvasir as a dataset is more complex and has a greater diversity of data. The performance of ResNet50 with an EFP layer showed greater improvements, though the relative increases compared to the baseline models were still evident. Specifically, ResNet50 + EFP achieved a 2.6% accuracy improvement over the ResNet50 baseline (84.57% vs. 81.93%), while reducing training time by more than 45 min. The ResNet18 with an EFP layer achieved respectable performance, with an accuracy of 77.41%, marginally above the ResNet18 baseline accuracy of 73.51%. The results of the study demonstrate an observable trend in the contribution made by including an EFP layer, improving training accuracy and training time, which enhances the performance of each model, particularly in more comprehensive datasets such as HyperKvasir.

Classifier	V1 (f1score)	V2 f1-score	$\phi_i = v1 - v2$	Rank	Signed Rank
LSVM	98.9	91.7	7.2	1	+1
QSVM	98.8	89.5	9.3	2	+2
CSVM	98.9	88.8	10.1	3	+3
MGSVM	98.7	87.3	11.4	6	+6
WNN	98.6	87.2	11.4	6	+6
MNN	98.3	87.0	11.3	5	+5
CGSVM	98.3	86.9	11.4	6	+6
BNN	98.3	86.5	11.8	8	+8
WKNN	96.8	86.2	10.6	4	+4
FKNN	96.7	83.1	13.6	9	+9

Table 14. Wilcoxon signed rank statistical analysis between the performance of classifiers on Kvasir v1 and Kvasir v2.

Method	Dataset	Year	Accuracy
Deep learning techniques[10]	Kvasir, CVC-ClinicDB and ETIS-LaribPolypDB	2019	96.5%.
Conventional techniques[13]	WCE	2020	98%
Deep learning[46]	Kvasir	2024	97%
Variational deep learning[47]	Kvasir-Capsule dataset	2024	87%
Dynamic multiclass learning[48]	Kvasir-Capsule and CAD-CAP	2024	96.47%, 96.72%
Our suggested approach	Kvasir	2024	99%

Table 15. Comparison with other state-of-the-art techniques.

Efficiency and real-time deployment

For the proposed gastrointestinal disease classification system to be deployed in real-time, several critical considerations must be addressed. The introduction of the EFP layer and the use of an NRMPO algorithm enable an increase in classification accuracy and robustness; however, this also requires optimization to perform at a computationally efficient level for real-world use. Furthermore, real-time prediction can be achieved by using lightweight versions of ResNet, and the model inference pipeline can utilize high-performance GPUs or edge devices with embedded AI accelerators. The separation of preprocessing, which includes top-bottom bias contrast enhancement and data augmentation, can be done in parallel, thereby reducing computing time. The proposed method described here is modularized to allow for integration into existing endoscopy imaging software; therefore, classification could occur automatically during an actual live procedure. However, further optimization and pruning of deep models would be necessary to provide clinicians with instantaneous feedback, supporting the faster and more accurate diagnosis of lesions while maintaining the standard operating clinical flow and duties intact and materially unchanged.

Wilcoxon signed rank test

In this section, we have implemented the Wilcoxon signed-rank test to assess the statistically significant differences among the classifiers on Kvasir v1 and Kvasir v2. Firstly, we defined the hypothesis as null hypothesis ($\aleph_0$), there is no significant difference between classifiers' performance on v1 and v2, and the Alternative hypothesis ($\aleph_L$) is the significant difference in classifiers' performance on v1 and v2 datasets. Initially, we measure the difference among the performance of Kvasir v1 and Kvasir v2 using $\phi_i = v1 - v2$, as shown in Table 14. After that, compute the absolute differences and rank them (smallest rank 1). After ranking, we assigned a sign to all the positive differences; therefore, all the ranks are in a positive sign, as shown in Table 14. After that, we compute the sum of signed ranks with the lesser sum using ω_+ and ω_-. The sum of ω_+ is 50 and $\omega_- = 0$. After that, $\min(\omega_+, \omega_-)$ is applied and got $\omega = 0$. In our case, the value of n=10 and confidence interval is 0.05, and $\omega_{crit} = 8$. For the decision, $\omega \leq \omega_{crit}$, $0 \leq 8$. Therefore, the null hypothesis $\aleph_0$ is rejected, meaning there is a statistically significant difference in classifier performances between the Kvasir v1 and Kvasir v2 datasets at a 0.05 confidence level, and the performance of Kvasir v1 is significantly better than that of Kvasir v2.

Comparison with other state of the art technique

We evaluate our approach, which extracts ulcer information from WCE images using CNNs. They utilized different datasets and achieved accuracies of 96.5% and 98%, respectively. Our method utilizes deep learning on the Kvasir V1 and V2 datasets, which comprise eight classes of gastrointestinal lesions. Our method achieves 99% accuracy and runs in 97.48 s. In terms of accuracy and speed, our technique surpasses the recent methods presented in Table 15.

Deep features extraction reasoning

ResNet architectures are typically executed in an end-to-end way to complete both feature extraction and classification in a single pipeline. While there is nothing inherently wrong with extracting features from modified ResNet representations, it was our purposeful choice to externally extract features using ResNet variants based on the challenges presented in our problem domain and research goals. In WCE, datasets are often small, unbalanced, and there is high inter-class similarity of visual information with slight differences in pathology. Performing end-to-end training in such spaces can lead to overfitting, which means not being able to utilize features fully.

To address these limitations, we introduced two key innovations:

- In ResNet architectures, the replacement of conventional pooling layers with our proposed Entropic Field Propagation (EFP) layer to better preserve fine-grained spatial information and enhance robustness against noise.
- The use of an external feature optimization stage, the Newton-Raphson-controlled Marine Predator Optimization, to refine and select only the most discriminative deep features before classification.

This method allowed us to separate the individual feature learning phase from the final classification phase of the analysis. It used multiple machine learning classifiers to be rigorous about the discriminative strength and generalization of the features learned. Furthermore, our studies on blending learned features and machine learning classification consistently outperformed the ResNet baseline, which was trained fully end-to-end for each dataset, and also reduced training time.

Thus, it should be emphasized that while the manual feature extraction step may appear to be a redundancy, it is a conscious design decision to implement novel feature selection and feature fusion techniques, It reduces overfitting in small, and imbalanced datasets, and facilitate interpretable evaluation of feature quality across classifiers, which may not be readily achievable in our application domain with pure end-to-end approaches. This methodological development is consistent with previous literature from the field of medical imaging, wherein additional performance gains were obtained through deep feature extraction with subsequent localized feature optimization and machine learning classification stages.

Conclusion

Our research suggests an automated technique for classifying stomach multiclass diseases using deep learning and the Newton-Raphson method. The fusion of top-down and bottom-up filtering increases the contrast of the primary capsule endoscopic image, which is beneficial for extracting useful features. The deep learning models are modified and trained on the Kvasir V1 and V2 enhanced datasets. In the testing phase, features are extracted from deep layers and optimized using the Newton-Raphson controlled MPA optimization algorithm. Optimal features are finally fused using a mean threshold-based fusion and classified using machine learning algorithms. This strategy reduces computing time while simultaneously improving accuracy. In WCE imaging datasets, the experimental approach was employed, resulting in increased accuracy.

Additionally, using an optimization strategy reduces the classification time. The limitation of this work is that ten classifiers are evaluated; the lack of end-to-end training across the whole system may limit the full potential of feature learning and optimization integration. In the future, we will consider the following issues.

(i) A single CNN Learning Model will be proposed and trained, replacing multiple Training Models, and an end-to-end trainable architecture will be explored.
(ii) Create a new CNN learning model and train it entirely from scratch for every stomach disease.
(iii) Instead of using vector length, which extends computing time, develop a new fusion strategy.

Data availability

This work's datasets are publicly available for research purposes. [https://datasets.simula.no/kvasir/](https:/datasets.simula.no/kvasir) .

Received: 22 May 2025; Accepted: 21 August 2025
Published online: 01 September 2025

References

1. Paul, J. Gastrointestinal tract infections, In: Disease Causing Microbes: Springer, 149–215. (2024).
2. Forman, D. & Burley, V. Gastric cancer: global pattern of the disease and an overview of environmental risk factors. *Best Pract. Res. Clin. Gastroenterol.* **20** (4), 633–649 (2006).
3. Sergi, C. M. Gastrointestinal tract, In: Pathology of Childhood and Adolescence: an Illustrated Guide: Springer, 255–424. (2020).
4. Naz, J. et al. Recognizing Gastrointestinal malignancies on WCE and CCE images by an ensemble of deep and handcrafted features with entropy and PCA based features optimization. *Neural Process. Lett.* **55** (1), 115–140 (2023).
5. Khan, M. A. et al. Gastrointestinal diseases recognition: a framework of deep neural network and improved moth-crow optimization with Dcca fusion. *Hum. -Cent Comput. Inf. Sci.* **12**, 25 (2022).
6. Collaborators, G. S. C. The global, regional, and National burden of stomach cancer in 195 countries, 1990–2017: a systematic analysis for the global burden of disease study 2017. *Lancet Gastroenterol. Hepatol.* **5** (1), 42 (2019).
7. Li, C., Li, L. & Shi, J. Gastrointestinal endoscopy in early diagnosis and treatment of Gastrointestinal tumors. *Pakistan J. Med. Sci.* **36** (2), 203 (2020).
8. Wang, S., Xing, Y., Zhang, L., Gao, H. & Zhang, H. Deep convolutional neural network for ulcer recognition in wireless capsule endoscopy: experimental feasibility and optimization, *Comput. Math. Methods Med.* **2019**(1), 7546215 (2019).
9. Sarfraz, M. S., Alhaisoni, M., Albesher, A. A., Wang, S. & Ashraf, I. StomachNet: optimal deep learning features fusion for stomach abnormalities classification. *IEEE Access.* **8**, 197969–197981 (2020).

10. Archana, R. & Jeevaraj, P. E. Deep learning models for digital image processing: a review. *Artif. Intell. Rev.* **57** (1), 11 (2024).
11. Obuchowicz, R., Strzelecki, M. & Piórkowski, A. *Clinical Applications of Artificial Intelligence in Medical Imaging and Image processing—A Review Vol*16p. 1870 (ed: MDPI, 2024).
12. Abhisheka, B., Biswas, S. K., Purkayastha, B., Das, D. & Escargueil, A. Recent trend in medical imaging modalities and their applications in disease diagnosis: A review. *Multimed. Tools Appl.* **83** (14), 43035–43070 (2024).
13. Xu, Y. et al. Advances in medical image segmentation: A comprehensive review of traditional, deep learning and hybrid approaches, *Bioengineering*. **11** (10) 1034, (2024).
14. Rahman, M. M., Wadud, M. A. H. & Hasan, M. M. Computerized classification of Gastrointestinal polyps using stacking ensemble of convolutional neural network. *Inf. Med. Unlocked*. **24**, 100603 (2021).
15. Rashid, M., Sharif, M., Raza, M., Sarfraz, M. M. & Afza, F. Object detection and classification: a joint selection and fusion strategy of deep convolutional neural network and SIFT point features. *Multimedia Tools Appl.* **78**, 15751–15777 (2019).
16. Gholami, E., Tabbakh, S. R. K. & Kheirabadi, M. Proposing method to Increase the detection accuracy of stomach cancer based on colour and lint features of tongue using CNN and SVM, *arXiv preprint arXiv*: **2020** 09962 (2011).
17. Khan, M. A. et al. Multiclass stomach diseases classification using deep learning features optimization, (2021).
18. Majid, A. et al. Classification of stomach infections: A paradigm of convolutional neural network along with classical features fusion and selection. *Microsc. Res. Tech.* **83** (5), 562–576 (2020).
19. Sharif, M., Akram, T., Yasmin, M. & Nayak, R. S. Stomach deformities recognition using rank-based deep features selection. *J. Med. Syst.* **43**, 1–15 (2019).
20. Akram, T., Sharif, M., Javed, K., Rashid, M. & Bukhari, S. A. C. An integrated framework of skin lesion detection and recognition through saliency method and optimal deep neural network features selection. *Neural Comput. Appl.* **32**, 15929–15948 (2020).
21. Nouri-Moghaddam, B., Ghazanfari, M. & Fathian, M. A novel multi-objective forest optimization algorithm for wrapper feature selection. *Expert Syst. Appl.* **175**, 114737 (2021).
22. Sun, Y. Review on computer vision in gastric cancer: Potential efficient tools for diagnosis, *arXiv preprint arXiv:2005.09459*, (2020).
23. Gholami, E. & Tabbakh, S. R. K. Increasing the accuracy in the diagnosis of stomach cancer based on color and Lint features of tongue. *Biomed. Signal Process. Control*. **69**, 102782 (2021).
24. Pannala, R. et al. Artificial intelligence in Gastrointestinal endoscopy, *VideoGIE*, **5** 598–613, (2020).
25. Guimarães, P., Keller, A., Fehlmann, T., Lammert, F. & Casper, M. Deep-learning based detection of gastric precancerous conditions, *Gut* **69** 4–6, (2020).
26. Lee, J. H. et al. Spotting malignancies from gastric endoscopic images using deep learning. *Surg. Endosc.* **33**, 3790–3797 (2019).
27. Mohammad, F. & Al-Razgan, M. Deep feature fusion and optimization-based approach for stomach disease classification, *Sensors* **22** no. 7, p. 2801, (2022).
28. Mehmood, A. et al. Prosperous human gait recognition: an end-to-end system based on pre-trained CNN features selection. *Multimed. Tools Appl.* **83** 1–21, (2020).
29. Bik, E. M. et al. Molecular analysis of the bacterial microbiota in the human stomach, *Proc. Nat. Acad. Sci.* **103** 732–737, (2006).
30. Thomas Abraham, J., Muralidhar, A., Sathyarajasekaran, K. & Ilakiyaselvan, N. A deep-learning approach for identifying and classifying digestive diseases, *Symmetry* **15** 379, (2023).
31. Pogorelov, K. et al. Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In *Proceedings of the 8th ACM on Multimedia Systems Conference*, pp. 164–169. (2017).
32. Hinata, M. & Ushiku, T. Detecting immunotherapy-sensitive subtype in gastric cancer using histologic image-based deep learning. *Sci. Rep.* **11** (1), 22636 (2021).
33. Bhardwaj, P., Kumar, S. & Kumar, Y. A comprehensive analysis of deep learning-based approaches for the prediction of gastrointestinal diseases using multi-class endoscopy images. *Arch. Comput. Methods Eng.* **30** 1–18, (2023).
34. Nouman Noor, M., Nazir, M., Khan, S. A., Song, O. Y. & Ashraf, I. Efficient gastrointestinal disease classification using pretrained deep convolutional neural network, *Electronics* **12** 1557, (2023).
35. Arowolo, M. O., Adebiyi, M. O., Nnodim, C. T., Abdulsalam, S. O. & Adebiyi, A. A. An adaptive genetic algorithm with recursive feature elimination approach for predicting malaria vector gene expression data classification using support vector machine kernels. *Walailak J. Sci. Technol. (WJST)*. **18** (17), 9849 (2021).
36. Naz, J. et al. A comparative analysis of optimization algorithms for gastrointestinal abnormalities recognition and classification based on ensemble XcepNet23 and ResNet18 features, *Biomedicines* **11** 1723, (2023).
37. Ahamed, M. F. et al. Automated detection of colorectal polyp utilizing deep learning methods with explainable AI. *IEEE Access*. **12** 78074–78100 (2024).
38. Ahamed, M. F. et al. Automated colorectal polyps detection from endoscopic images using multiresunet framework with attention guided segmentation. *Human-Centric Intell. Syst.* **4** (2), 299–315 (2024).
39. Ahamed, M. F. et al. Detection of various gastrointestinal tract diseases through a deep learning method with ensemble ELM and explainable AI. *Expert Syst. Appl.* **256**, 124908 (2024).
40. Ahamed, M. F., Shafi, F. B., Nahiduzzaman, M., Ayari, M. A. & Khandakar, A. Interpretable deep learning architecture for Gastrointestinal disease detection: A Tri-stage approach with PCA and XAI. *Comput. Biol. Med.* **185**, 109503 (2025).
41. Smedsrud, P. H. et al. Kvasir-Capsule, a video capsule endoscopy dataset. *Sci. Data*. **8** (1), 142 (2021).
42. Perez, L. & Wang, J. The effectiveness of data augmentation in image classification using deep learning, *arXiv preprint arXiv:1712.04621*, (2017).
43. Karras, T. et al. Training generative adversarial networks with limited data. *Adv. Neural. Inf. Process. Syst.* **33**, 12104–12114 (2020).
44. Liu, X., Wang, C., Bai, J. & Liao, G. Fine-tuning pre-trained convolutional neural networks for gastric precancerous disease classification on magnification narrow-band imaging images, *Neurocomputing*, **392** 253–267, (2020).
45. Faramarzi, A., Heidarinejad, M., Mirjalili, S. & Gandomi, A. H. Marine predators algorithm: A nature-inspired metaheuristic. *Expert Syst. Appl.* **152**, 113377 (2020).
46. Niharika, N., Raju, S. S., Akash, R., Hussain, I. & Victor, K. J. An in-depth analysis of Gastrointestinal abnormalities using endoscopic imaging and deep learning. In Challenges in Information, Communication and Computing Technology: CRC, 202–207. (2025).
47. Singh, B., Kumar, P. & Jain, S. K. Combining the variational and deep learning techniques for classification of video capsule endoscopic images. *J. Imag. Inf. Med.* 1–20, (2025).
48. Li, X., Wu, Q. & Wu, K. Wireless capsule endoscopy anomaly classification via dynamic multi-task learning. *Biomed. Signal Process. Control*. **100**, 107081 (2025).

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00218176) and the Soonchunhyang University Research Fund. This work was supported through Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R410), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author contributions

The authors Saddaf Rubab, Muhammad Jamshed, Muhammad Attique Khan, and Nouf Abdullah Almujally contributed to methodology, software, conceptual design, and original write-up. The authors Robertas Damaševičius, Amir Hussain, Neunggyu Han, and Yunyoung Nam contributed to funding, supervision, review write-up, and project administration.

Funding

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00218176) and the Soonchunhyang University Research Fund. This work was supported through Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R410),Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Declarations

Competing interests

The authors declare that they have no conflicts of interest to report regarding the present study.

Additional information

Correspondence and requests for materials should be addressed to S.R. or Y.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.